

## TOPOLOGY OF LARGE-SCALE STRUCTURES OF GALAXIES IN TWO DIMENSIONS – SYSTEMATIC EFFECTS

STEPHEN APPLEBY<sup>a</sup>, CHANGBOM PARK<sup>a</sup>, SUNGWOOK E. HONG<sup>a</sup>, AND JUHAN KIM<sup>b</sup><sup>a</sup>School of Physics, Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul 02455, Korea and<sup>b</sup>Center for Advanced Computation, Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul 02455, Korea*Draft version February 16, 2017*

## ABSTRACT

We study the two-dimensional topology of the galactic distribution when projected onto two-dimensional spherical shells. Using the latest Horizon Run 4 simulation data, we construct the genus of the two-dimensional field and consider how this statistic is affected by late-time nonlinear effects – principally gravitational collapse and redshift space distortion (RSD). We also consider systematic and numerical artifacts such as shot noise, galaxy bias, and finite pixel effects. We model the systematics using a Hermite polynomial expansion and perform a comprehensive analysis of known effects on the two-dimensional genus, with a view toward using the statistic for cosmological parameter estimation. We find that the finite pixel effect is dominated by an amplitude drop and can be made less than 1% by adopting pixels smaller than 1/3 of the angular smoothing length. Nonlinear gravitational evolution introduces time-dependent coefficients of the zeroth, first, and second Hermite polynomials, but the genus amplitude changes by less than 1% between  $z = 1$  and  $z = 0$  for smoothing scales  $R_G > 9\text{Mpc}/h$ . Non-zero terms are measured up to third order in the Hermite polynomial expansion when studying RSD. Differences in shapes of the genus curves in real and redshift space are small when we adopt thick redshift shells, but the amplitude change remains a significant  $\sim \mathcal{O}(10\%)$  effect. The combined effects of galaxy biasing and shot noise produce systematic effects up to the second Hermite polynomial. It is shown that, when sampling, the use of galaxy mass cuts significantly reduces the effect of shot noise relative to random sampling.

## 1. INTRODUCTION

The distribution of galaxies in the universe contains a wealth of information. Their clustering properties can be used for cosmological parameter estimation and to test the fundamental assumptions upon which the standard cosmological model is based. Conventional applications of galaxy data in cosmology have focused on the  $N$ -point correlation functions – effectively measuring the clustering as a function of scale.

In parallel to the standard  $N$ -point analysis, an alternative approach has been developed that uses the topology of the galaxy distribution for cosmological parameter estimation (Park & Kim 2010; Zunckel et al. 2011; Speare et al. 2015; Blake et al. 2014) and also searching for non-Gaussian and modified gravity signals (Matsubara & Jain 2001; Hikage et al. 2006; Wang et al. 2012; James 2012). In this approach, the genus of the galactic density field is obtained from the data, and the conservation of genus with redshift provides a standard ruler or population by which one can reconstruct the ‘true’ cosmology. The two methods extract similar information (both involve measuring the power spectrum); however, their sensitivity to systematic effects such as bias and redshift space distortions (RSDs) differ. In addition, the genus is sensitive to the phase structure of the density field on small scales, hence further information can be extracted from the genus curve by considering the evolution of its shape with redshift. This requires modeling of the genus shape in terms of cumulants of the density field in the nonlinear regime (Pogosyan et al. 2009; Gay et al. 2012; Codis et al. 2013).

The genus is perhaps the simplest and most commonly utilized measure of topology. It is a member of the Minkowski Functionals – a set of scalar quantities that form a complete basis of valuations of a space (Hadwiger 1957). Their origin can be traced to early works on integral geometry, and they have been utilized in cosmology for nearly three decades (Gott et al. 1986, 1988; Ryden et al. 1989; Gott et al. 1988; Melott et al. 1989; Park et al. 1992; Park & Gott 1991; Matsubara 1994, 1996; Schmalzing et al. 1996; Park et al. 2005). In this paper, we focus exclusively on the two-dimensional genus of large-scale structure, which has been studied extensively in the literature (Melott et al. 1989; Coles & Plienis 1991; Park et al. 1992; Coles et al. 1993; Colley et al. 2000; Park et al. 2001; Hoyle et al. 2002; Colley 1997; Gott et al. 2007; Wang et al. 2015).

The matter density field is three-dimensional in nature. However, in practice we do not always have access to full three-dimensional spatial information, for a number of reasons. One reason is that we measure the density field in redshift space, which is biased relative to its real space counterpart, due to both linear and nonlinear effects. In addition, cataloging large numbers of galaxies is typically achieved using photometric bands to estimate their redshift. Although recent advances in this field are impressive (Bilicki et al. 2013; Laigle et al. 2016), photometric redshift estimation remains a dominant source of uncertainty.

In this work, we study the two-dimensional genus of the galaxy distribution, tomographically binning galaxies in shells. By taking sufficiently thick redshift slices, we aim to mitigate both RSD and photometric redshift effects. The price we pay for these advantages is that the two-

dimensional genus contains less information, relative to its three-dimensional counterpart. The Fourier modes parallel to the line of sight are effectively averaged in each redshift bin.

We use state-of-the-art simulations (Kim et al. 2015) to study the large-scale topological properties of the matter density field, using the distribution of dark matter particles and mock galaxy catalogs as tracers. The goal is a comprehensive study of the genus curve and how its properties are modified by various nonlinear and numerical processes. Many of these effects have been studied previously, both analytically and numerically: finite pixel effects have been expounded in Melott et al. (1989), and gravitational evolution and RSD have been studied in Matsubara & Suto (1996) and Matsubara (1996, 2000); Park et al. (2005). This work is a companion paper to Kim et al. (2014), where a similar analysis was performed for the three-dimensional genus.

The paper will proceed as follows. In section 2, we define the two-dimensional genus and calculate it for a simple Gaussian field. Following this, we describe how the genus is modified by various late-time effects, and use the latest Horizon Run 4 (Kim et al. 2015) N-body simulation to calculate departures from Gaussianity. We briefly describe the simulation data used in the analysis and proceed to study the consequences of pixel size, gravitational evolution, RSD, and shot noise in sections 4.1-4.4. We summarize in section 5.

## 2. TWO-DIMENSIONAL GENUS

Our aim is to study the two-dimensional genus of slices of the matter density field, using both dark matter particles and simulated galaxies as tracers. We begin with a discussion of the genus for a Gaussian random field  $\delta(x, y, z)$  in three-dimensional space, with power spectrum  $P_{3D}(k_x, k_y, k_z) = P_{3D}(k_{3D})$  where  $k_{3D} = \sqrt{k_x^2 + k_y^2 + k_z^2}$ . We are interested in two-dimensional slices of this field, so we define a line of sight (without loss of generality  $z$ ) and introduce a top-hat window function along this axis. The resultant two-dimensional field is then defined as the integral

$$\hat{\delta}_{2D}(x, y) = \int \delta(x, y, z) F_z(z) dz, \quad (1)$$

with top hat

$$F_z(z) = \frac{1}{\Delta_z} [\Theta(z - z_{\min}) - \Theta(z - z_{\max})], \quad (2)$$

and bin thickness  $\Delta_z = z_{\max} - z_{\min}$  along the line of sight.  $\Theta(z)$  is the Heaviside step function.

We also smooth the field in the  $(x, y)$  plane perpendicular to the line of sight - for this purpose we adopt a Gaussian smoothing kernel. The final density field  $\delta_{2D}$  is given by

$$\delta_{2D}(x, y) = \int \delta(\tilde{x}, \tilde{y}, \tilde{z}) F_z(\tilde{z}) d^2 \tilde{r} d\tilde{z}, \quad (3)$$

where  $\mathbf{r} = (x, y)$  is the position vector in the two-dimensional space and

$$F_2D(|\mathbf{r}|) = \frac{1}{2\pi R_G^2} \exp\left[-\frac{r^2}{2R_G^2}\right]. \quad (4)$$

where  $R_G$  is the two-dimensional Gaussian smoothing scale. The original three-dimensional field has power spectrum  $P_{3D}(k_{3D})$ , and the two-dimensional  $P_{2D}(k_{2D})$  is related to  $P_{3D}(k_{3D})$  as

$$P_{2D}(k_{2D}) = \frac{L}{2\pi} \int P_{3D}[(k_{2D}^2 + k_z^2)^{1/2}] \exp[-k_{2D}^2 R_G^2] \times \frac{\sin^2(k_z \Delta_z)}{k_z^2 \Delta_z^2} dk_z, \quad (5)$$

where  $L^3$  is the total volume under consideration and  $k_{2D} = \sqrt{k_x^2 + k_y^2}$ . For a Gaussian field, one can analytically calculate the expectation value of the genus curve, which is given by Melott et al. (1989)

$$G_{2D}(\nu) = \frac{1}{(2\pi)^{3/2}} \frac{\langle k_{2D}^2 \rangle}{2} \nu \exp[-\nu^2/2], \quad (6)$$

where  $\langle k_{2D}^2 \rangle = \int k_{\perp}^2 P_{2D}(k_{\perp}) d^2 k_{\perp} / \int P_{2D}(k_{\perp}) d^2 k_{\perp}$  is the square of the wavenumber  $k_{2D}$  averaged over the two-dimensional power spectrum  $P_{2D}(k_{2D})$ . Thus, in the case of a Gaussian field, the amplitude of the genus is related to the shape of the power spectrum in the vicinity of the two-dimensional smoothing scale  $R_G$ , and scales with the inverse square mean separation between peaks. The amplitude is a function of both  $R_G$  and  $\Delta_z$ .

The quantity  $\nu$  is the density threshold used to define the excursion set for which we calculate the genus, normalized by the variance of the field. However, in what follows we choose not to calculate the genus directly in terms of the threshold overdensity  $\nu$ . Instead, we relate this quantity to the area fraction of the field that lies above a particular value. We define  $\nu_A$  as

$$f_A = \frac{1}{\sqrt{2\pi}} \int_{\nu_A}^{\infty} \exp[-t^2/2] dt, \quad (7)$$

where  $f_A$  is the fractional area of the field above  $\nu_A$ . For example, when  $\nu_A = 0$ , we calculate the genus of the density field that occupies exactly half of the total area of the two-dimensional space, regardless of whether it is Gaussian-distributed. In the case of a Gaussian field,  $\nu_A$  is identical to  $\nu$  within statistical fluctuations. More generally, there exists a bijective map between  $\nu$  and  $\nu_A$ : however, the  $\nu_A$  parameterization has been used to separately study non-Gaussianity beyond that in the one-point function (Gott et al. 1987; Weinberg et al. 1987; Melott et al. 1988).

From an observational perspective, the galaxy samples that we typically use for cosmology are not homogeneous – we introduce radial selection functions, and also line-of-sight effects such as RSDs must also be accounted for. Therefore, it will be convenient to project the density field onto shells centered on our location, and to calculate the genus on this surface rather than flat two-dimensional planes. Our analysis will closely follow that of Schmalzing & Gorski (1998), whose method we review for completeness.

For a two-dimensional field on an arbitrary surface  $Q$ , we can write the genus as Gott et al. (1989)

$$G_{2D} = \frac{1}{2\pi} \int_{\partial Q} \kappa d\ell, \quad (8)$$

where  $d\ell$  denotes the line element along the surface of  $Q$  and  $\kappa$  is the geodesic curvature. We calculate the surface density of the genus (that is, the genus per unit area) on the unit sphere, which is given by (Schmalzing & Gorski 1998)

$$g_{2D}(\nu) = \frac{1}{4\pi} \int_Q da \tilde{\delta}(\delta - \nu) \frac{2\delta_{;1}\delta_{;2}\delta_{;12} - \delta_{;1}^2\delta_{;22} - \delta_{;2}^2\delta_{;11}}{\delta_{;1}^2 + \delta_{;2}^2}, \quad (9)$$

where  $da$  now represents a surface element of the 2-space  $Q$ , and covariant derivatives (first and second) are denoted with a semicolon. The indices 1, 2 represent an arbitrary coordinate system covering  $Q$  and  $\tilde{\delta}(x)$  denotes the Dirac delta function.

To calculate the genus, our first step will be to pixelate the sphere. When doing so, we must calculate the first and second derivatives of  $\delta$  at each pixel center. We also approximate the delta function as

$$\tilde{\delta}(\delta - \nu) \simeq \frac{1}{\epsilon} \mathbf{I}_{(-\epsilon/2, +\epsilon/2)}(\delta - \nu), \quad (10)$$

where  $\mathbf{I}_A(x) = 1$  for  $x \in A$  and  $\mathbf{I}_A(x) = 0$  otherwise. The value of  $\epsilon$  is chosen to be small, and the final result should be independent of its value. However, taking  $\epsilon$  too small creates a sparse set of pixels that satisfy  $\delta - \nu \in (-\epsilon/2, +\epsilon/2)$ , generating numerical noise. If  $\epsilon$  is too large, then the delta function is not well-approximated by (10). Based on numerical testing of the stability of the genus, we take  $\epsilon = 2.0 \times 10^{-3}(\delta_{\max} - \delta_{\min})$  in what follows, where  $\delta_{\min, \max}$  are the minimum/maximum values of the density field on the sphere.

The discretized equivalent of equation (9) is

$$g_{2D} \simeq \frac{1}{N_{\text{pix}}} \sum_{i=1}^{N_{\text{pix}}} \tilde{\delta}^{(i)} \frac{2\delta_{;1}^{(i)}\delta_{;2}^{(i)}\delta_{;12}^{(i)} - (\delta_{;1}^{(i)})^2\delta_{;22}^{(i)} - (\delta_{;2}^{(i)})^2\delta_{;11}^{(i)}}{(\delta_{;1}^{(i)})^2 + (\delta_{;2}^{(i)})^2}, \quad (11)$$

where  $\delta_{;1}^{(i)}$  (for example) represents a discrete approximation to the full covariant derivative and  $N_{\text{pix}}$  is the total number of pixels. On non-flat  $Q$ -surfaces, it is practical to calculate covariant derivatives in Fourier space. We use the standard angular coordinate system

$$ds_2^2 = d\theta^2 + \sin^2(\theta)d\phi^2. \quad (12)$$

Coordinate labels (1, 2) in equations (9) and (11) correspond to  $\theta, \phi$ .

For a Gaussian field, we can relate the amplitude of the genus to the angular power spectrum  $\mathcal{C}_\ell$  as Schmalzing & Gorski (1998)

$$g_{2D} = \frac{1}{2(2\pi)^{3/2}} \frac{\sigma_1^2}{\sigma_0^2} \nu \exp[-\nu^2/2], \quad (13)$$

$$\sigma_0^2 = \sum_{\ell=1}^{\infty} (2\ell + 1) \mathcal{C}_\ell, \quad (14)$$

$$\sigma_1^2 = \sum_{\ell=1}^{\infty} (2\ell + 1) \ell(\ell + 1) \mathcal{C}_\ell. \quad (15)$$

In Figure 1 (top panel), we exhibit the genus curve for

a Gaussian realization of a field drawn from a  $\Lambda$ CDM power spectrum. The blue points are the values calculated using equation (11), and the black solid line is the theoretical expectation value equation (13). The bottom panel exhibits the angular power spectrum of the Gaussian field on the unit sphere. The shape of the genus curve is completely fixed, and only the amplitude carries information.

From inspection of equations (13 – 15), it is clear that the amplitude of the genus is a measure of ratios of cumulants of the density field. As we are measuring ratios of the power spectrum, the amplitude will be insensitive to both linear bias and the linear growth rate. Despite this, one can still extract cosmological information from the genus amplitude by comparing its value in different redshift shells. Specifically, one can use the fact that the genus of the dark matter density field should be a conserved quantity when smoothed on large scales. However, if we choose an incorrect cosmology, then the physical smoothing scale that we adopt will systematically evolve with redshift, along with the measured surface area of the shell. The genus is sensitive to both of these quantities, and the two effects (incorrect smoothing scale and surface area) will not exactly cancel because the field is not scale-invariant. Hence, by measuring the genus in different redshift shells, we can reconstruct cosmological parameters by minimizing the evolution of the genus.

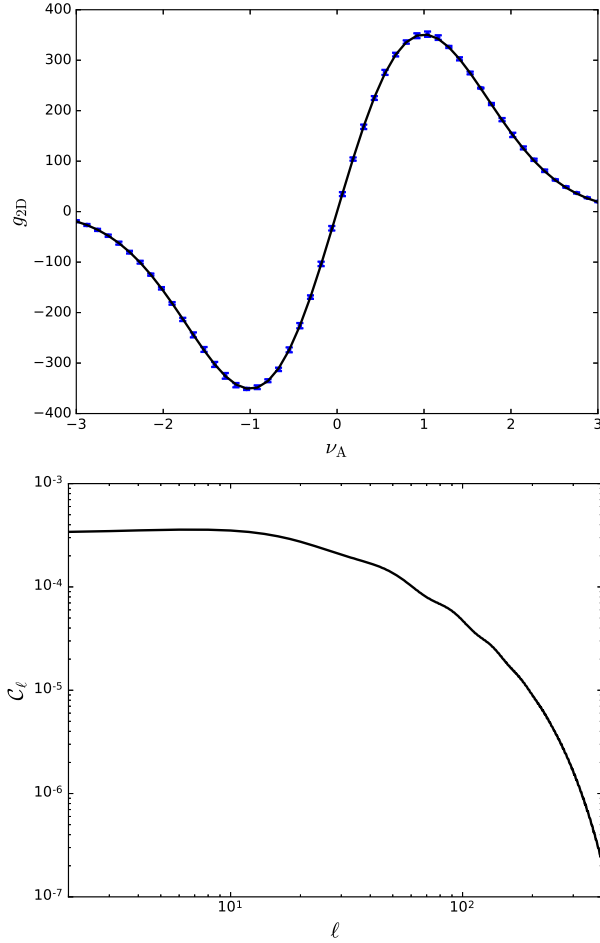
### 3. GENUS OF A NON-LINEAR DENSITY FIELD – HORIZON RUN 4

Having completed our discussion on Gaussian random fields, we now consider a discrete point distribution of dark matter obtained from the Horizon Run 4 (HR4) simulation. The density field, obtained by smoothing the point distribution, is now non-Gaussian due to a number of nonlinear effects - chiefly nonlinear gravitational clustering. Before continuing, we briefly describe the simulation.

Horizon Run 4 (Kim et al. 2015) is the latest data release from the Horizon Run project<sup>1</sup>. It is a dense, cosmological-scale  $N$ -body simulation that gravitationally evolved  $N = 6300^3$  particles in a  $V = (3150 \text{ Mpc}/h)^3$  volume box. The simulation uses a modified version of GOTPM code and initial conditions are calculated using second-order Lagrangian perturbation theory, taking into account the experiments on the impact of initial conditions by L'Huillier et al. (2014). The cosmological parameters used are given in Table 3, and details of the simulation can be found in Kim et al. (2015). We will predominantly make use of snapshot data at  $z = 0, 0.3, 0.7, 1.0, 4.0$ . In each snapshot box, we generate spheres of radius  $d_{\text{cm}} \sim 738 \text{ Mpc}/h$ . With this choice, we can embed  $n = 8$  non-overlapping shells into the simulation box.

We study two distinct density fields constructed from different point catalogs - the dark matter particle data and a mock galaxy catalog constructed in Hong et al. (2016). We thin the particle data by a factor of four in each dimension - for our choice of pixel size the resulting number density is sufficient to minimize shot noise. The thinning procedure involves selecting every fourth particle along each dimension in the Lagrangian coordi-

<sup>1</sup> <http://sdss.kias.re.kr/astro/Horizon-Runs>



**Figure 1.** A genus curve for a Gaussian random field drawn from a sample angular power spectrum  $C_\ell$ , exhibited in the lower panel. The blue points and error bars were obtained by generating  $N = 200$  realizations of a Gaussian random field on the sphere and using our pixelated method to reconstruct the genus. The solid black line is the analytic prediction, equation (13).

mates  $(i, j, k)$ , which outputs  $(6300/4)^3$  particles. Here,  $i, j, k$  are integer values. The initial particles are positioned at each grid point in the Lagrangian coordinates and they are perturbed to construct the desired density fluctuations, which are consistent with the input linear matter power spectrum. Therefore, to select fractions of particles from the total number, we have used their Lagrangian positions. Mock galaxies are assigned by the most bound halo particle-galaxy correspondence scheme. Survival time of satellite galaxies after merger is calculated by adopting the merger timescale model described in Jiang et al. (2008).

As we are now considering a non-Gaussian field, one can expect the genus curve to change in both amplitude and shape. In the mildly nonlinear regime, nonlinear effects have been analytically calculated and their effect on the genus shape is well-described by a low-order Hermite polynomial expansion (Matsubara 2000). In the following sections, we typically calculate the residual of two genus curves based upon different assumptions - we define this quantity as  $\Delta g_{2D}$ . The quantity  $\Delta g_{2D}$  will be defined differently in each of the following sections, according to the specific systematic quantity that we wish

to study. For example, in section 3.2 we analyze departures of the genus curve from its Gaussian limit; in this case, we define  $\Delta g_{2D}$  as the difference between the measured genus curve and its Gaussian expectation (13). In section 4.2 we study the evolution of the genus with redshift; in that case  $\Delta g_{2D}$  will denote the difference between the genus curve at high and low- $z$ . We clearly define  $\Delta g_{2D}$  at the beginning of each section.

Theoretically, we model  $\Delta \hat{g}_{2D}$  using a Hermite polynomial expansion

$$\Delta \hat{g}_{2D}(\nu_A, a_{0-4}) = A \exp[-\nu_A^2/2] \sum_{i=0}^4 a_i H_i(\nu_A), \quad (16)$$

with constant  $a_i$ , where  $A$  is the amplitude calculated from equation (13)

$$A = \frac{1}{2(2\pi)^{3/2}} \frac{\sigma_1^2}{\sigma_0^2}, \quad (17)$$

and the Hermite polynomials are given by

$$H_0(x) = 1, \quad (18)$$

$$H_1(x) = x, \quad (19)$$

$$H_2(x) = x^2 - 1, \quad (20)$$

$$H_3(x) = x^3 - 3x, \quad (21)$$

$$H_4(x) = x^4 - 6x^2 + 3. \quad (22)$$

We calculate  $\Delta g_{2D}$  from the data for a set of  $n = 8$  independent shells within the Horizon Run box and then perform a  $\chi^2$  minimization of

$$\chi^2 = \sum_{i=1}^{n_g} \frac{[\Delta g_{g,2D}(\nu_{A,i}) - \Delta \hat{g}_{2D}(\nu_{A,i}, a_{0-4})]^2}{\epsilon_i^2}, \quad (23)$$

to find the best fit  $a_{0-4}$  values. Here,  $i$  subscripts denote the  $n_g = 50$  linearly spaced values of  $\nu_A$  at which we calculate the genus  $g_{2D,i}$ , and  $\epsilon_i^2$  is the population mean uncertainty of  $\Delta g_{2D}(\nu_{A,i})$ , calculated using the eight distinct shells in the HR4 box. Specifically,  $\epsilon_i^2$  is obtained using a simple student t-test as  $\epsilon_i = t_{\alpha/2} s_i / \sqrt{n}$ , where  $s_i$  is the standard deviation of  $\Delta g_{2D}(\nu_{A,i})$ ,  $n = 8$ , and the  $\alpha$  subscript denotes the confidence limit chosen. We adopt the 95% percentile for  $t_{\alpha/2}$ . We use the package COSMOMC (Lewis 2013; Lewis & Bridle 2002) as a generic sampler for the minimization. In applying this minimization procedure, we have assumed that the noise is Gaussian distributed, which is an increasingly poor approximation in the rare event tails of the distribution. When applying the genus statistic to data, accurate theoretical modeling of the statistical uncertainty must be undertaken. This will be the subject of future work.

### 3.1. Results

We select eight locations in the  $z = 0$  snapshot data as observer points, separated such that we measure non-overlapping regions, and then construct shells between two redshift limits  $z_{\min} = 0.25$  and  $z_{\max} = 0.27$  (that is, shells of width  $\Delta = 0.02$ ). Here,  $z_{\min, \max}$  are redshifts

| Parameter          | Fiducial Value    |
|--------------------|-------------------|
| $\Omega_{m0}$      | 0.26              |
| $\Omega_{\Lambda}$ | 0.74              |
| $n_s$              | 0.96              |
| $\sigma_8$         | 0.794             |
| $\Delta$           | 0.02              |
| $\theta_G$         | 27.5              |
| $z_{\min}$         | 0.25              |
| $N_s$              | 1024              |
| $N_{\text{pix}}$   | $12 \times N_s^2$ |

**Table 1**

Fiducial parameters used in the Horizon Run 4 simulation, and the parameters used to calculate the genus in this work.  $\Delta$  is the redshift thickness of the spherical shells,  $\theta_G$  the angular smoothing scale (in arcminutes) applied to the unit sphere.  $z_{\min}$  is the redshift that defines the inner boundary of the shell,  $N_{\text{pix}}$  is the number of pixels used to discretize the sphere.

relative to the observer, and  $z_{\text{mid}} = (z_{\min} + z_{\max})/2$  denotes the approximate redshift of the center of the shell. Because we use snapshot data,  $(z_{\min}, z_{\max})$  are not true redshifts. However, the  $z = 0$  snapshot that we predominantly use is a reasonable approximation to lightcone data for small  $z$ .

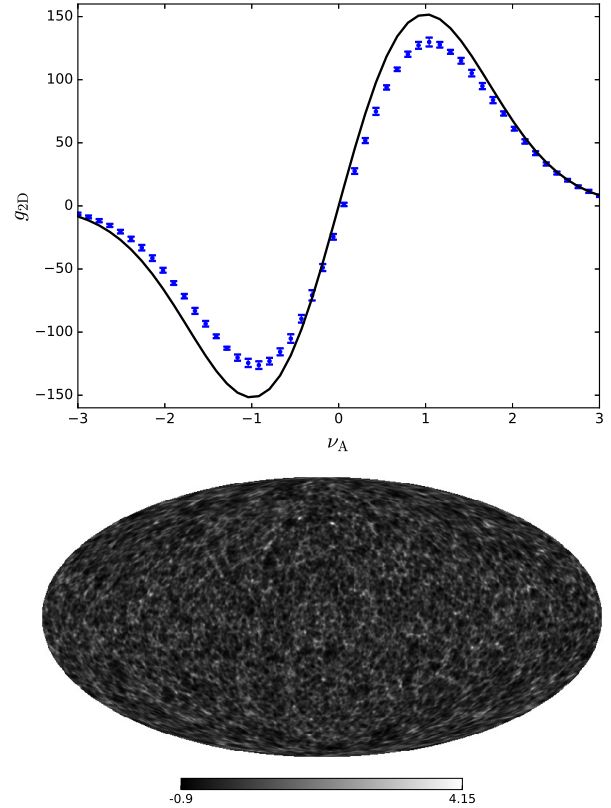
We fix the total number of pixels as  $N_{\text{pix}} = 12 \times N_s^2$ , where  $N_s = 1024$ , and pixelate the sphere according to the HEALPix<sup>2</sup> equal area scheme. The sphere is initially decomposed into twelve curvilinear quadrilaterals, and then subsequent subdivisions are made by dividing each pixel into four new ones until the desired pixel resolution is achieved. Details of the procedure can be found in Gorski et al. (2005).

We use the HR4 cosmological parameters to relate the comoving distance along the line of sight  $d_{\text{cm}}(z)$  and the redshift  $z$ , and then bin all simulated galaxies that lie within the redshift range  $(z_{\min}, z_{\max})$  according to their nearest pixel center. Note that, when we perform the same calculation using galaxy data, no cosmological information is required at this stage. Cosmological dependence will only arise when we wish to compare the genus in different redshift shells, or when we wish to relate redshift and real space quantities.

After binning the galaxies or dark matter particles, we obtain a series of pixels containing  $n_i$  points, where  $i$  is the pixel identifier. We calculate the average  $\bar{n}$  over the entire pixel range and define  $\delta_i = (n_i - \bar{n})/\bar{n}$ . We then perform two-dimensional Gaussian smoothing of  $\delta_i$  on the unit sphere using angular scale  $\theta_G$ , which is related to the physical scale at the shell location as  $R_G \simeq d_{\text{cm}}(z_{\text{mid}})\theta_G$ . We typically take  $\theta_G = 27.5$  arcmin in what follows, corresponding to a physical scale of  $R_G = 5.9$  Mpc/h.

To calculate the covariant derivatives of the field, we use HEALPix to first Fourier transform the field, and then calculate the derivatives using recursion relations between the Legendre polynomials.

Our baseline parameters are  $\theta_G = 27.5$  arcmin and  $\Delta = 0.02$ ; however, we will vary these in the following section. We exhibit the dark matter genus curve as a function of  $\nu_A$  in Figure 2. The blue points and error bars represent the mean and standard deviation of the eight samples, and the black solid line denotes the genus



**Figure 2.** (Top panel) The dark matter genus curve as a function of  $\nu_A$ . The points and error bars represent the sample mean and standard deviation of eight distinct shells generated on the  $z = 0$  HR4 snapshot data. The black solid line is the Gaussian expectation value based on the angular power spectrum of the two-dimensional density field. (Bottom panel) A Mollweide projected density map of a shell of the Horizon Run 4 dark matter density field, smoothed with our fiducial parameters  $\Delta = 0.02$ ,  $\theta_G = 27.5$  arcmin.

curve constructed from the Gaussian expectation value (13) using the angular power spectrum calculated from the data.

The discrepancy between the Gaussian curve and the data indicates that our fiducial parameters are probing the nonlinear regime. The most striking departure from Gaussianity is an amplitude decrease, known as gravitational smoothing (Melott et al. 1989). However, the shape is also modified in an asymmetric manner. We discuss these effects further in the following section.

### 3.2. Dependence on smoothing scales $\Delta$ and $\theta_G$

Before studying gravitational and systematic effects, we begin by considering the dependence of the genus on the two smoothing scales  $\Delta$  and  $\theta_G$ . The dependence on  $\theta_G$  is crucial for cosmological parameter estimation, and the shell thickness  $\Delta$  can also potentially contain cosmological information.

In the Gaussian limit, the genus amplitude will depend quadratically on  $\theta_G$ . Although we are considering the nonlinear regime, we still expect growth of the amplitude as we decrease the smoothing scale  $\theta_G$ , by virtue of the increasing number of structures being resolved.

Increasing  $\Delta$  will ‘Gaussianize’ the field along the line of sight, in the sense that when  $\Delta$  is larger than the typical scale of bound structures we begin to stack structures along the line of sight. Stacking Fourier modes with cor-

<sup>2</sup> <http://healpix.sourceforge.net>

related phases (as is the case in the late Universe) will have the effect of randomizing the phases. Mathematically, one can simply state that, as we increase  $\Delta$ , we are increasing the size of the real space top hat function. In Fourier space, we approach a delta function kernel when integrating over the modes parallel to the line of sight. All information in these modes is lost in the thick shell limit.

To examine sensitivity of  $g_{2D}$  to  $(\theta_G, \Delta)$ , we define the function

$$\Delta g_{2D} = g_{2D,HR4}(\nu_A) - g_{2D,G}(\nu_A), \quad (24)$$

where  $g_{2D,HR4}(\nu_A)$  is the genus extracted from the HR4 data and  $g_{2D,G}(\nu_A)$  is the Gaussian expectation value (13), and the angular power spectrum is constructed from the data. We use both dark matter particle and mock galaxy data in this section, and apply the expansion (16) to (24). The resulting coefficients  $a_{0-4}$  represent amplitude and shape departures from Gaussianity.

We first fix  $\Delta = 0.02$  and vary  $\theta_G$ , and then fix  $\theta_G = 27.5$  arcmin and vary  $\Delta$ . The amplitude coefficient  $a_1$  is exhibited as a function of  $\theta_G$  and  $\Delta$  in Figure 3. The blue triangles/solid line represent the amplitude obtained from the dark matter data, and the yellow squares/dashed line represents the mock galaxy catalog. The  $a_1$  parameter is negative for the entire  $(\theta_G, \Delta)$  range considered - this is due to gravitational smoothing (Melott et al. 1989; Park & Gott 1991; Park et al. 2005). The genus amplitude of the matter density field exhibits a stronger shift from the Gaussian limit than the galaxy distribution. This amplitude drop is not related to the linear bias between the samples - in fact, the genus amplitude is insensitive to (constant) bias factors. We note the approach to Gaussianity  $a_1 \rightarrow 0$  with increasing smoothing scale  $\theta_G$ .

When we fix  $\theta_G = 27.5$  arcmin and vary  $\Delta$  over the range  $0.0025 < \Delta < 0.025$ , we also find a slow approach to Gaussianity  $a_1 \rightarrow 0$  with increasing  $\Delta$ . This is consistent with our expectations when stacking modes along the line of sight. When considering photometric galaxy data, we must fix the width of the tomographic bins such that they are wider than the typical photometric redshift uncertainty of the catalog. Current state-of-the-art in the field will force us to choose  $\Delta > 0.02$ , which corresponds to a physical scale of  $\sim 60$  Mpc/h. Even on these large scales, the genus can exhibit departures from Gaussianity at low redshift.

We plot the shape parameters  $a_{0,2,3,4}$  as a function of  $\theta_G$  (top panel,  $\Delta = 0.02$ ) and  $\Delta$  (bottom panel,  $\theta_G = 27.5$  arcmin) in Figure 4. Once again, the solid/dashed lines represent dark matter and mock galaxy samples respectively. The  $a_{0,2,3,4}$  parameters describe departures of the shape of the genus curve away from its Gaussian form.

One can see that the genus curve is essentially insensitive to  $a_{3,4}$  even on very small angular scales. The only significant shape modifiers are related to the shift parameter  $a_0$  and the quadratic term  $a_2$ . This is in agreement with semi-analytic results (Matsubara 2000) when one adopts the  $\nu_A$  parameterization. The  $a_2$  term changes sign when we take thin slices of the field. For negative  $a_2 < 0$ , the genus exhibits a so-called ‘meatball shift’ to the left of the origin  $\nu_A = 0$ . This effect was observed in

(Melott et al. 1989) and is reproduced here for  $\Delta < 0.005$ . For thick slices, in which structures are stacked, one finds  $a_2 > 0$  which corresponds to a shift of the genus curve to the right of the origin. This suggests a preponderance of ‘bubbles’ in the density field. The scale at which this transition occurs will depend on the correlation length of the point distribution. For the dark matter field, the transition occurs at smaller  $\Delta$ .

For thick slices the parameters  $a_{0,2,3,4}$  remain relatively insensitive to  $\Delta$ . We also observe a slow decrease in  $a_{0,2}$  as we increase the angular smoothing scale  $\theta_G$ . This corresponds to an approach to the Gaussian limit. This occurs for both dark matter and simulated galaxies. The dark matter data exhibits stronger non-Gaussian behavior than the galaxies at all scales probed.

The statistic (24) informs us of departures of the measured genus from Gaussianity. To study only the effect of  $\theta_G, \Delta$  on the genus curve, we define the function

$$R = \frac{a_1(\theta_G, \Delta)}{a_1(\theta_0, \Delta_0)}, \quad (25)$$

where  $a_1$  is the amplitude of the genus and  $\theta_0, \Delta_0$  are arbitrary values of our smoothing parameters that conveniently normalize  $R$ . We calculate  $R$  for both a Gaussian field and the Horizon Run 4 dark matter data by fixing  $\Delta_0 = 0.005$ ,  $\theta_0 = 13.75$  arcmin and varying  $\theta_G$ , and subsequently  $\Delta$ . We exhibit  $R$  in Figure 5 as a function of  $\theta_G$  (top panel) and  $\Delta$  (bottom panel). We calculate its value for both a Gaussian field (blue solid line) and the HR4  $z = 0$  dark matter field (yellow dashed line). One can observe the expected  $R \sim \theta_G^{-2}$  behavior for the Gaussian field, and a less-steep dependence for the nonlinear data. The dark matter field is quite insensitive to  $\Delta$  in the thick shell limit  $\Delta > 0.015$ .

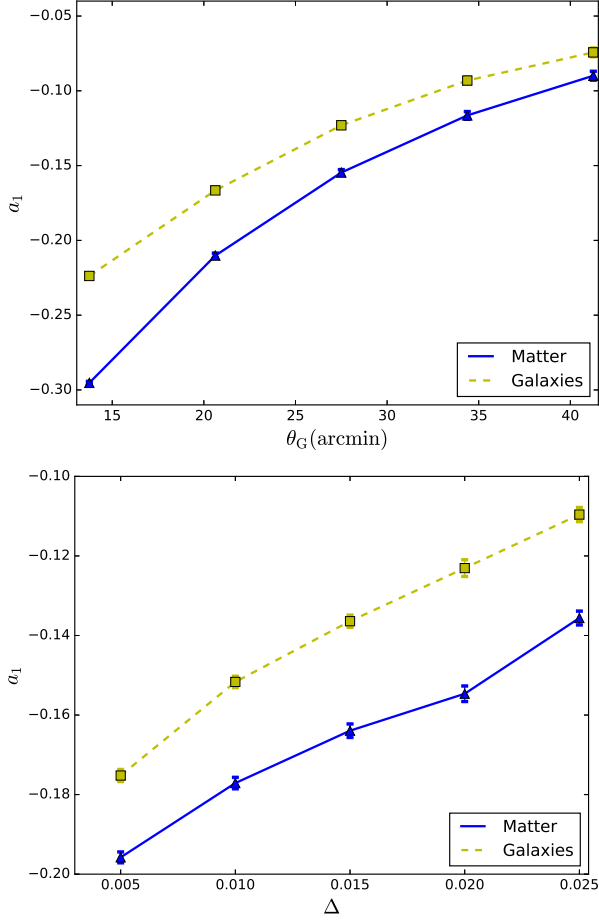
#### 4. NON-LINEAR EFFECTS

The genus of a Gaussian density field has a particularly simple form, and all information is encoded in the amplitude of the function. However, the matter density in the low redshift universe is highly nonlinear and non-Gaussian. Numerous complications arise that can modify both the shape and amplitude of the genus curve. In what follows, we discuss four dominant effects: finite size pixels, gravitational evolution, RSD and shot noise/galaxy bias.

##### 4.1. Pixel Effects

The first departure from Gaussianity considered is an entirely numerical artifact. A pixelated density field  $\delta_{ijk}$  will perfectly represent its continuous counterpart  $\delta(x, y, z)$  only in the limit  $p \rightarrow 0$ , where  $p$  is the typical size of a pixel. Pixels of finite size will introduce spurious modifications to the genus of order  $\mathcal{O}(p^2/k^2)$ . Analytic forms exist that correct the genus for finite pixel effects (Melott et al. 1989); however, the form of the correction depends on both the shape of the pixel and the mass binning scheme. Here, we fit the effect using the Hermite expansion (16). We take a set of smoothing scales  $\theta_G$  and for each calculate the genus using three different pixel sizes  $N_s = (1024, 512, 256)$ , where  $N_{\text{pixel}} = 12 \times N_s^2$ . We then define

$$\Delta g_{2D,\text{pix}} = g_{2D}(\theta_G, N_s) - g_{2D}(\theta_G, N_s = 1024), \quad (26)$$



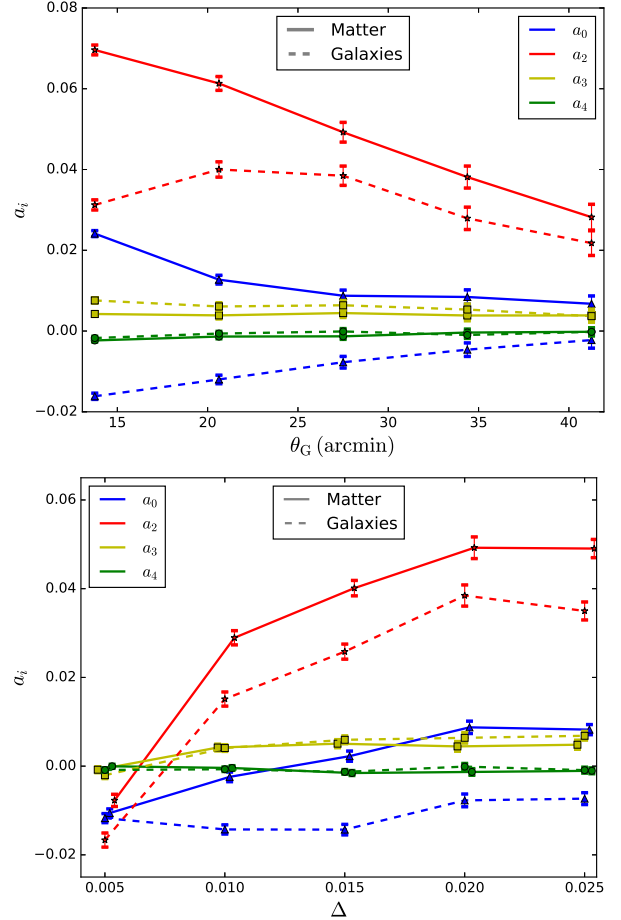
**Figure 3.** (Top Panel) The departure of the genus amplitude from its Gaussian limit ( $a_1$ ) as a function of  $\theta_G$  in arcminutes, for our eight fiducial shells in the  $z = 0$  snapshot box. The blue triangles denote dark matter and yellow squares the simulated galaxy catalog. (Bottom Panel) The same quantity as a function of the shell thickness  $\Delta$ . The amplitude of the simulated density field is lower than the Gaussian limit, a well-known effect. One finds a slow approach to Gaussianity with increasing  $\Delta$  and  $\theta_G$ .

where  $g_{2D}(\theta_G, N_s)$  are the values of the genus extracted from the data for  $N_s = (512, 256)$ . We fit the expansion (16) to (26) - If we assume that the finest pixelation of the sphere,  $N_s = 1024$ , provides an accurate representation of the true density field, then the coefficients  $a_{0-4}$  will inform us of the dependence of the genus curve on the size of the pixels.

In Figure 6 (top panel) we exhibit the coefficient  $a_1$  as a function of  $p/\theta_G$ . We can consider this curve as representative of the magnitude of the finite pixel effect on the amplitude of the genus as a function of  $p/\theta_G$ . The shape of the pixels in the HEALPix projection are not standardized, so we define  $p = (10800/\pi)\sqrt{4\pi/N_{\text{pixel}}}$  (in arcminutes). We use the mock galaxy data in this section.

The term  $a_1$  decreases with  $p/\theta_G$ . The pixel effect is  $\sim 1.2\%$  when  $p/\theta_G \sim 0.5$  and  $\sim 3\%$  at  $p/\theta_G \sim 1$ . One should choose  $p < \theta_G/3$  to safely ensure that pixelation effects are less than 1%.

The Hermite polynomial coefficients  $a_{0,2}$  are exhibited in the bottom panel of Figure 6. These quantities do not vary appreciably with pixel size for  $p/\theta_G < 1$ , although there is increasing scatter for large pixels. For  $p < \theta_G/3$ ,



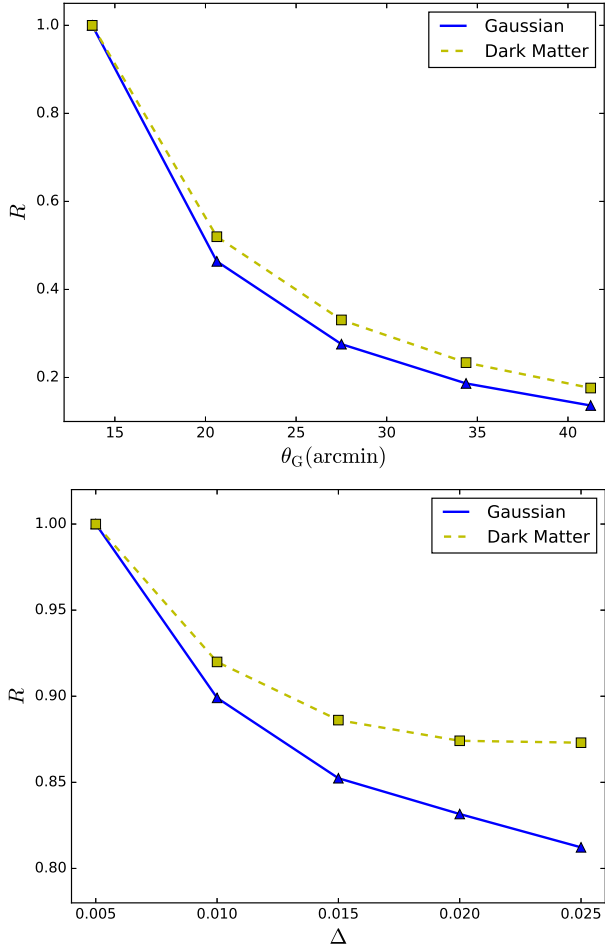
**Figure 4.** (Top Panel) The shape parameters  $a_{0,2,3,4}$  as a function of  $\theta_G$  in arcminutes, for our fiducial redshift shells  $z_{\text{min}} = 0.25$  and  $\Delta = 0.02$ . We exhibit  $a_{0,2,3,4}$  as blue triangles, red stars, yellow squares, and green circles respectively. Solid/dashed lines denote the dark matter/galaxy data. The  $a_{0,2}$  terms are the most significant contributors to the distortion from a purely Gaussian genus curve, and  $a_{3,4}$  constitute negligible effects. (Bottom Panel) the same quantities  $a_{0,2,3,4}$  now as a function of  $\Delta$ , fixing  $\theta_G = 27.5$  arcmin. The same labeling scheme applies. One can see that the shape parameters are relatively insensitive to  $\Delta$  for  $\Delta > 0.015$ , and  $a_{0,2}$  are the dominant contributors to departures from the Gaussian form. There is a sign change in  $a_2$  as we transition to smaller scales.

no spurious shape dependence will arise due to finite size pixels.

#### 4.2. Gravitational Clustering

As the density field evolves from high redshift, it undergoes collapse into knots and filaments. The process of gravitational evolution will significantly modify the morphology of structures (on small scales and at late times, where the field is strongly nonlinear). The effect of gravitational evolution on the power spectrum of the dark matter density field is to transfer power from large to small scales. As the amplitude of the genus is related to an integral with the power spectrum acting as a window function (see equation (13)), one might expect the amplitude to increase from high to low redshift. However, the effect of gravitational smoothing will decrease the amplitude of a gravitationally evolved field relative to a Gaussian field drawn from the same power spectrum. Furthermore, when averaging the density field over suit-



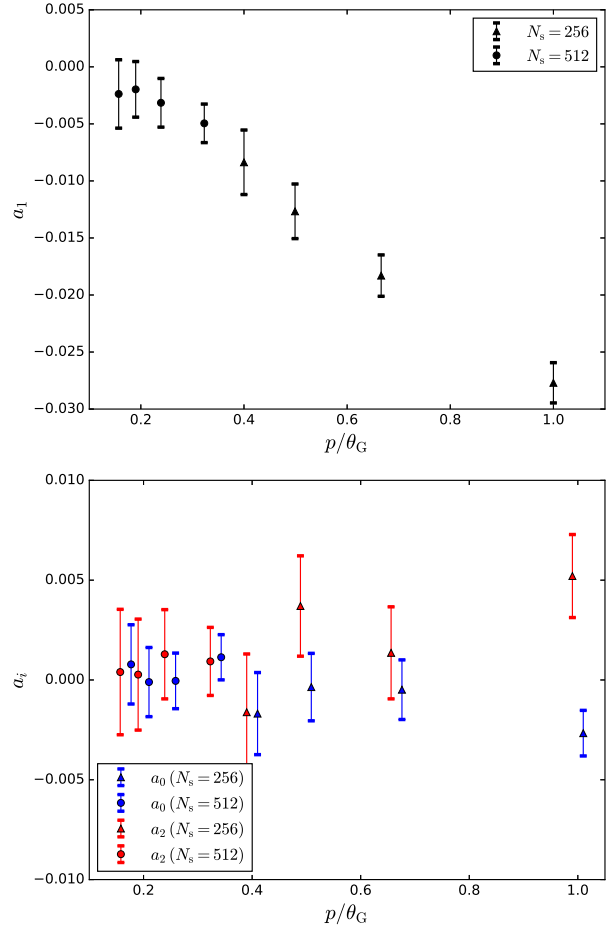


**Figure 5.** (Top Panel) The statistic  $R = a_1(\theta_G, \Delta = 0.002)/a_1(\theta_G = 13.75 \text{ arcmin}, \Delta = 0.002)$  as a function of  $\theta_G$  for a Gaussian field (blue solid) and the HR4  $z = 0$  dark matter field (yellow dashed). The Gaussian case decreases as  $\theta_G^{-2}$  as expected; however the nonlinear field exhibits a more shallow dependence on the smoothing scale. (Bottom panel) The  $R$  statistic as a function of  $\Delta$  for a Gaussian field (blue solid) and the HR4  $z = 0$  dark matter field (yellow dashed). Once again, the nonlinear field exhibits less-steep dependence on the smoothing scale. The  $\Delta$  dependence becomes increasingly negligible in the thick shell limit  $\Delta > 0.015$ .

ably large scales, the amplitude should exhibit no appreciable evolution, as the linear power spectrum preserves its shape with time and the number of structures will be conserved. In this section, we use Horizon Run 4 dark matter data to study the evolution of the genus with redshift.

We take the dark matter snapshot data at five redshifts  $z = (0, 0.3, 0.7, 1.0, 4.0)$  and calculate the genus. Specifically, we place eight observers in the simulation box at each epoch and calculate the genus curve in shells located at distance  $d_{\text{cm}} = 738 \text{ Mpc/h}$  from the observer. We fix  $(\theta_G, \Delta)$  to their fiducial values in this section.

Our comparison of snapshot data does not accurately reflect the realities of observational cosmology, where one would use lightcone data and galaxies as opposed to dark matter. However, in this work we wish to disentangle various systematics that act to modify the genus curve. As such, we measure the evolution of the genus using dark matter particles; in doing so we largely avoid shot noise and galaxy bias that are present in galaxy catalogs.



**Figure 6.** (Top panel) The Hermite polynomial coefficient  $a_1$  as a function of pixel size  $p/\theta_G$ . Here  $a_1$  describes the finite pixel effect on the amplitude of the genus, as explained in the text. The black circles have been generated using pixel size  $N_s = 512$  and black triangles  $N_s = 256$ . One can observe a systematic decrease in the amplitude as we increase the pixel size. Pixel effects are sub-percent level for  $p < \theta_G/3$ . (Bottom panel) The Hermite coefficient parameters  $a_0, a_2$  (blue and red points respectively), as a function of pixel size  $p/\theta_G$ . Triangles and circles represent  $N_s = 256$  and  $N_s = 512$ . The  $a_{0,2}$  parameters are consistent with zero over the range considered, indicating that pixel size effects do not significantly modify the shape of the genus curve.

We consider these systematics in section 4.4.

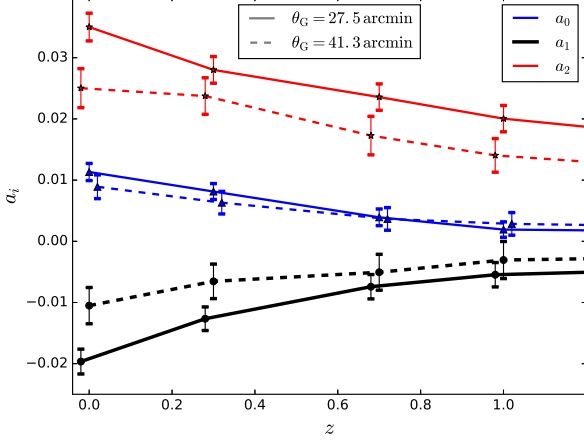
After calculating the genus in eight shells at each redshift, we construct the difference

$$\Delta g_{2\text{D,grav}}(\nu_A, z) = g_{2\text{D}}(\nu_A, z) - g_{2\text{D}}(\nu_A, z_0), \quad (27)$$

and fit the low-order Hermite polynomial approximation (16) to this function. Here,  $z_0$  is some arbitrary high redshift, that we take as  $z_0 = 4$ . The coefficients  $a_{0-4}$  will inform us of the evolution of the genus amplitude ( $a_1$ ) and shape ( $a_{0,2,3,4}$ ) with redshift. We exhibit the coefficients  $a_{0-2}(z)$  in Figure 7. As in Figure 4,  $a_{0,2}$  are represented as blue triangles and red stars respectively, and now black lines/points correspond to  $a_1$ . We omit  $a_{3,4}$ , which are consistent with zero at all redshifts. Solid (dashed) lines denote the choice of smoothing parameters  $\theta_G = 27.5, (41.3) \text{ arcmin}$  respectively, with fixed  $\Delta = 0.02$ .

When we adopt small-scale smoothing  $\theta_G =$



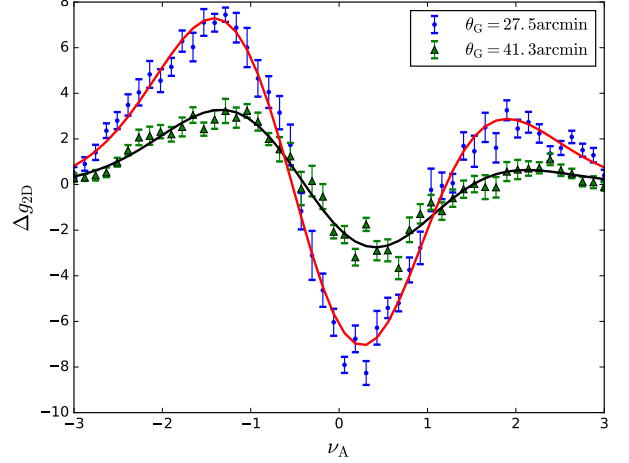


**Figure 7.** The evolution of the Hermite coefficients  $a_{0-2}$  with redshift. The solid (dashed) lines represent smoothing scales  $\theta_G = 27.5, (48.1)$  arcmin respectively. The genus amplitude, represented by  $a_1$ , decreases as redshift decreases due to gravitational smoothing - a Gaussian field will maximize the genus amplitude of a given power spectrum. This effect is mitigated as we increase the smoothing scale. The shape parameters  $a_{0,2}$  also exhibit considerable redshift evolution, and all quantities approach zero at early times, indicating an approach to Gaussianity.

27.5 arcmin, one can observe a monotonic increase in the parameters  $a_{0,2}$  with decreasing redshift, indicating an increasingly non-Gaussian field at late times. The amplitude decreases slightly over the range  $0 < z < 1$ , this is due to gravitational smoothing. This effect is mitigated when we use a larger smoothing scale  $\theta_G = 41.3$  arcmin, and the change in the amplitude of the genus over the range  $0 < z < 1$  is less than  $\sim 1\%$ .

In a companion paper, we will use the amplitude of the genus for cosmological parameter estimation, which requires the genus amplitude to act as a standard ruler. This condition is satisfied only when the density field is smoothed over scales sufficiently large that the number of structures is conserved or when the systematic evolution is well-understood. In Figure 7 one can observe clear evolution in the Hermite coefficients  $a_{0,1,2}$ , which is mitigated when we smooth over larger scales (dashed curves). Because we are probing the nonlinear regime, the parameter  $a_1$  does not completely specify the form of the genus. Both  $a_0$  and  $a_2$  are non-zero and evolve over the range  $z \sim (0, 1)$ , suggesting an evolving shape. The shape evolution will be asymmetric around  $\nu_A = 0$ , as the  $H_{0,2}(\nu_A)$  Hermite polynomials do not respect the fact that the Gaussian genus is an odd function. The  $a_{0,2}$  contributions are necessarily redshift-dependent, and should approach their Gaussian expectation value  $a_{0,2} = 0$  at high redshift and/or large smoothing scales  $\theta_G$ . The shape parameters  $a_0, a_2$  grow with the variance of the field  $\sigma_0$ .

We exhibit the asymmetry of  $\Delta g_{2D, \text{grav}}(z = 0)$  for the parameter choices  $\theta_G = 27.5, 41.3$  arcmin in Figure 8 (blue and green points), and the best fit theoretical curves are also shown as solid red and black lines, respectively. The error bars represent the 68% uncertainty on the population mean of  $\Delta g_{2D, \text{grav}}$  based on the  $n = 8$  shells within the simulation box. For  $\theta_G = 41.3$  arcmin,  $\Delta g_{2D, \text{grav}}$  is more symmetric about  $\nu_A = 0$  but as we smooth over successively smaller scales the under-dense region  $\nu_A < 0$  exhibits an excess in the genus curve rel-



**Figure 8.** The quantity  $\Delta g_{2D, \text{grav}}(z, \nu_A)$  at  $z = 0$  for two different angular smoothing scales  $\theta_G = 27.5, 41.3$  arcmin (blue and green points). The error bars denote the  $\sim 68\%$  uncertainty on the population mean, based on our  $n = 8$  distinct shells obtained from the simulation box. One observes increasingly nonlinear and asymmetric behavior as we decrease the smoothing scale, with nonlinear effects more apparent at low density. The red and black curves are the best fit to the data sets when we adopt the Hermite expansion (16). The curve shape is largely dictated by the asymmetric  $a_{0,2}$  parameters.

ative to  $\nu_A > 0$ , in agreement with three-dimensional results of Kim et al. (2014). In Hoyle et al. (2002), a preponderance of isolated clusters relative to isolated voids has been observed in galaxy data. On the other hand, according to Park et al. (2005) the shift and asymmetry of the genus curve depend on the smoothing scale and redshift for a given cosmology, as well on the kind of density tracer adopted.

In Matsubara (2003), the non-Gaussian signal in the two-dimensional genus was calculated to linear order in  $\sigma_0$

$$g_{2D}(\nu_A) = A e^{-\nu^2/2} \left[ H_1(\nu_A) + \left[ \frac{2}{3} (S^{(1)} - S^{(0)}) H_2(\nu_A) + \frac{1}{3} (S^{(2)} - S^{(0)}) \right] \sigma_0 + \mathcal{O}(\sigma_0^2) \right], \quad (28)$$

where the skewness parameters  $S^{(0)}, S^{(1)}, S^{(2)}$  are related to the three point statistics of the density field as

$$S^{(0)} = \frac{\langle \delta^3 \rangle}{\sigma_0^4}, \quad (29)$$

$$S^{(1)} = -\frac{3}{4} \frac{\langle \delta^2 (\nabla^2 \delta) \rangle}{\sigma_0^2 \sigma_1^2}, \quad (30)$$

$$S^{(2)} = -3 \frac{\langle (\nabla \delta \cdot \nabla \delta) (\nabla^2 \delta) \rangle}{\sigma_1^4}. \quad (31)$$

We can define  $\hat{a}_{0,2}$  coefficients as

$$\hat{a}_0 \equiv \frac{1}{3} (S^{(2)} - S^{(0)}) \sigma_0, \quad (32)$$

$$\hat{a}_2 \equiv \frac{2}{3} (S^{(1)} - S^{(0)}) \sigma_0. \quad (33)$$

The expansion (28) qualitatively agrees with our numerical analysis; throughout this work the  $H_0(\nu_A)$  and  $H_2(\nu_A)$  Hermite polynomials dominate the non-Gaussian signal. One can test more rigorously the domain in which the expansion (28) matches our numerical results by estimating the three point functions  $S^{(0)}$ ,  $S^{(1)}$ ,  $S^{(2)}$  for the density field and directly comparing  $\hat{a}_{0,2}$  and  $a_{0,2}$ .

We measure the skewness parameters  $S^{(0)}$ ,  $S^{(1)}$ ,  $S^{(2)}$  from the data. Because we measure both  $a_{0,2}$  and  $S^{(0)}$ ,  $S^{(1)}$ ,  $S^{(2)}$  from the same simulation snapshot data, we can expect the two approaches to produce statistically consistent  $a_0 \sim \hat{a}_0$ ,  $a_2 \sim \hat{a}_2$  coefficients as long as we are within the domain of applicability of the small  $\sigma_0$  expansion. When the expansion breaks down, we expect to find significant disagreement between the parameters.

In figure 9 we exhibit the  $a_0, \hat{a}_0$  (top panel) and  $a_2, \hat{a}_2$  (bottom panel) parameters as a function of  $\sigma_0$ . We repeat our analysis using three snapshot boxes  $z = (0, 0.7, 4)$  and various smoothing scales. Here, circular points are the  $a_{0,2}$  parameters obtained by fitting the expansion (16) directly to the genus curves, and squares are calculated using the skewness parameters  $S^{(0)}$ ,  $S^{(1)}$ ,  $S^{(2)}$  in the definitions (32, 33). The magenta, yellow, and black points correspond to redshift boxes  $z = (4, 0.7, 0)$  respectively. One can observe agreement between the two methods for small  $\sigma_0$  in both parameters. The  $z = 4$  high-redshift field exhibits close agreement between the two methods, with a breakdown at large  $\sigma_0 > 0.2$  for the low-redshift samples. The  $H_0$  coefficient fails more severely at  $\sigma_0 \sim 0.2$  than  $H_2$ . The  $H_0$  coefficient is a function of  $S^{(2)}$ , which is particularly sensitive to small-scale physics, and thus is the most difficult skewness parameter of the three to measure as it involves a higher-order derivatives of the density field, cf. equation (31). The  $a_2$  parameter is well-measured even at relatively large  $\sigma_0$  values  $\sigma_0 \sim 0.25$ .

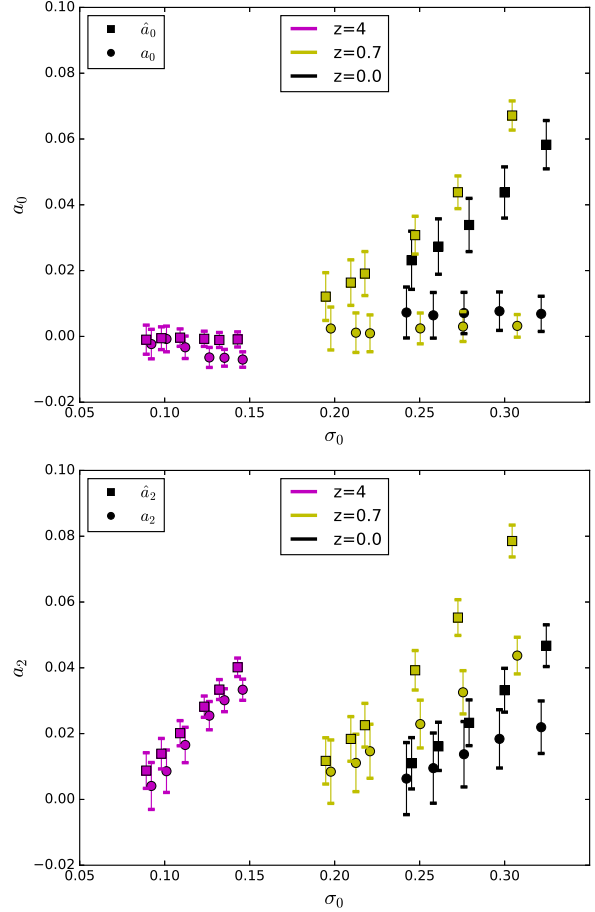
We conclude that the genus amplitude does not vary appreciably when the density field is smoothed over sufficiently large scales  $R_G > 9\text{Mpc}/h$ , and furthermore that the genus shape parameters provide an unbiased estimate of the third order cumulants of the density field for moderate  $\sigma_0$  values  $\sigma_0 < 0.2$ .

#### 4.3. Redshift Space Distortion

Galaxies possess a peculiar velocity along the line of sight at which we observe them. These motions cause the observed redshift of a galaxy to misrepresent its distance from the observer. The density field in redshift space is therefore distorted relative to its real space equivalent - on large scales, coherent in-fall produces pancake-like distortions, whereas on small scales peculiar velocities of bound objects produce the so-called ‘finger of God’ effect along the line of sight. The peculiar velocities are themselves generated by the gravitational field, leading to a complicated dynamical system.

Thus far, we have largely focused on dark matter. In this section we choose to study RSD effects on the HR4 simulated galaxy catalog, so that we can apply our results to observational data in future work.

The effect of peculiar velocities on the mock galaxy catalog is straightforward to calculate, as we know both the real and redshift space position of every data point. We therefore calculate the two-dimensional genus in both



**Figure 9.** (Top panel) The  $H_0$  Hermite polynomial coefficient obtained by fitting the expansion (16) directly to the measured genus curves (circular points) and the theoretical prediction (32) (squares), as a function of  $\sigma_0$ . We repeat our analysis for three redshift shells  $z = (4, 0.7, 0)$  (magenta, yellow and black points respectively). The agreement between  $a_0$  and  $\hat{a}_0$  is reasonable for small  $\sigma_0$ , with a clear breakdown of the second-order perturbation theory for  $\sigma_0 > 0.2$ . (Bottom panel) As in the top panel, but now the parameter  $H_2$  Hermite polynomial coefficient. One can observe agreement between  $a_2$  and  $\hat{a}_2$  to  $\sigma_0 \sim 0.25$ , followed by a similar (but less severe) breakdown in the expansion.

spaces, and compare them to leading-order theoretical predictions. In real space we simply take a random point in the snapshot box as an observer, and calculate the comoving distance from the observer to each galaxy. We then bin the galaxies into shells and calculate the genus according to section 2. To obtain the density field in redshift space, we first calculate the redshift corresponding to the comoving distance between galaxy and observer using the correct cosmology. We label this redshift  $z_c$ . We then calculate the peculiar velocity of the galaxy along the line of sight relative to the observer, which we define as  $v_{\text{los}}$ . The actual redshift that would be measured in a galaxy survey is given by

$$1 + z_{\text{obs}} = (1 + z_c) \left( 1 + \frac{v_{\text{los}}}{c} \right), \quad (34)$$

where  $c$  is the speed of light. The redshift space distance to the galaxies is then defined as  $d_{\text{rsd}}(z_{\text{obs}})$ . We use  $z_{\text{obs}}$  to calculate the two-dimensional genus of spherical shells in redshift space.

In Matsubara (1996), the effect of RSD on the two-dimensional genus was calculated using linear theory. For a flat two-dimensional slice, if we define the angle between the plane and the line of sight as  $\theta_S$  then the redshift and real space genus are related as

$$g_{2D}^{\text{RSD}}(\nu, \theta_S) = A_{\text{RSD}} g_{2D}^{\text{real}}(\nu), \quad (35)$$

where

$$A_{\text{RSD}} = \frac{3}{2} \sqrt{\left(1 - \frac{C_1}{C_0}\right) \left[1 - \frac{C_1}{C_0} + \left(\frac{3C_1}{C_0} - 1\right) \cos^2(\theta_S)\right]}, \quad (36)$$

and

$$\frac{C_1}{C_0} = \frac{1}{3} \frac{1 + (6/5)fb^{-1} + (3/7)(fb^{-1})^2}{1 + (2/3)fb^{-1} + (1/5)(fb^{-1})^2}. \quad (37)$$

Here,  $b$  is the bias factor,  $f = \dot{D}/(HD) \simeq \Omega_{\text{mat}}^{6/11}$ , and  $D$  is the linear growth rate. This result is valid to linear order in the density perturbation  $\delta$ . One can observe that the effect of RSD on the genus simply corresponds to an amplitude shift, as long as we restrict our analysis to suitably large scales. This calculation is valid for flat slices of the density field; however we expect it to provide a reasonable approximation to our numerical results, as we are choosing thick redshift slices  $\Delta = 0.02$  and the physical smoothing scale  $R_G$  on the sphere is much smaller than the observer-shell distance.

We exhibit the genus curves in real (blue error bars) and redshift (green error bars) space in Figure 10. One can observe the amplitude drop in redshift space. If we intend to use the statistic for cosmological parameter estimation, then it is important that we correct for this effect.

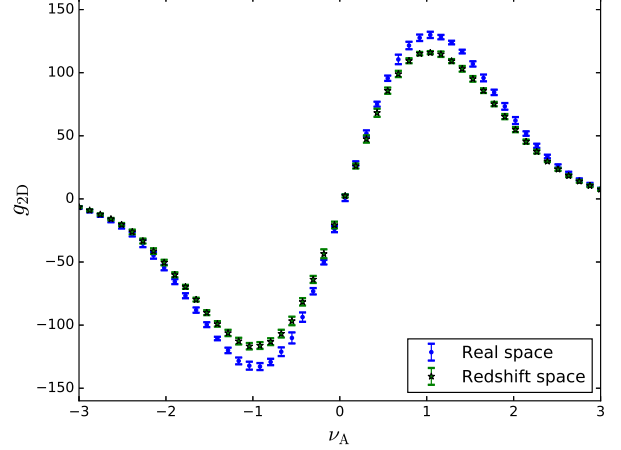
We define a new residual  $\Delta g_{2D, \text{RSD}}$  as the difference between the redshift and real space genus curves -

$$\Delta g_{2D, \text{RSD}}(\nu_A) = g_{2D, \text{RSD}}(\nu_A) - g_{2D, \text{real}}(\nu_A), \quad (38)$$

where  $g_{2D, \text{RSD}}(\nu_A, \theta_G, \Delta)$ ,  $g_{2D, \text{real}}(\nu_A, \theta_G, \Delta)$  are the genus curves extracted from the  $z = 0$  snapshot HR4 mock galaxy data in redshift and real space respectively. We fit the Hermite expansion (16) to (38). The Hermite coefficient  $a_1$  will describe the amplitude shift and  $a_{0,2,3,4}$  the shape modification as we move between real and redshift space. In the linearized limit,  $a_{0,2,3,4}$  should be consistent with zero. Based on our experience in section 3.2, we expect both  $g_{2D, \text{RSD}}(\nu_A)$  and  $g_{2D, \text{real}}(\nu_A)$  will depart from the Gaussian limit (13). However we make no assumptions regarding the Gaussianity of the field in this section.

Redshift space effects will be sensitive to both  $\theta_G$  and  $\Delta$ . For this reason, we repeat our calculation for two different smoothing scales  $\theta_G = (27.5, 48.1)$  arcmin and two different shell thickness values  $\Delta = (0.005, 0.02)$  (which we call thin and thick shells respectively).

In figure 11 (top panel), we exhibit the amplitude coefficient  $a_1$  for the four possible combinations of smoothing scales  $(\theta_G, \Delta)$ . The linear theory prediction  $A_{\text{RSD}} - 1$  is shown as a solid vertical line, and the error bars are the uncertainty on the mean given our sample of  $n = 8$  shells.



**Figure 10.** Genus curves in real (blue points) and redshift (green stars) space. One can observe a decrease in the amplitude in redshift space, as predicted from linear theory. Error bars denote the standard deviation of the measurement of eight distinct shells in the snapshot data, and we use mock galaxy data.

One can observe that the linear prediction is in approximate agreement with our numerical results, to within  $\sim 1\%$  when we use  $\theta_G = 48.1$  arcmin smoothing. We have found in previous sections that the genus curve amplitude  $a_1$  departs significantly from its Gaussian form. Despite this, the amplitude shift between real and redshift space is roughly consistent with the linear prediction (36). However, to use this statistic for cosmological parameter estimation we must improve this agreement to the sub-percent level.

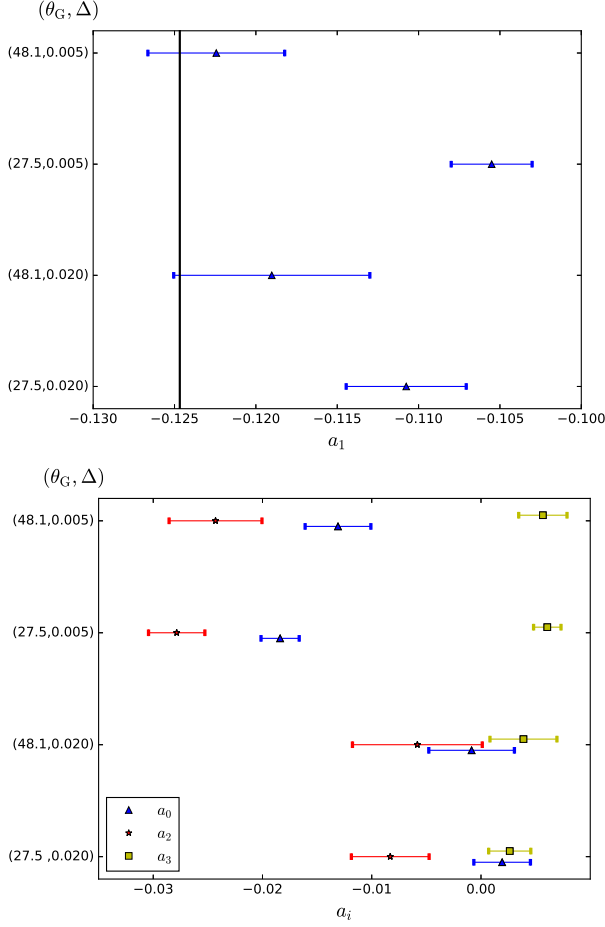
When using a large smoothing scale  $\theta_G = 48.1$  arcmin, the amplitude drop  $a_1$  is statistically consistent for both thick  $\Delta = 0.02$  and thin  $\Delta = 0.005$  slices. This is not the case as we decrease the smoothing scale on the sphere to  $\theta_G = 27.5$  arcmin, at which point non-linear effects should be accurately modeled. The  $\theta_G = 27.5$  arcmin smoothing cases produce an amplitude shift that is lower than anticipated from linear theory; however, the difference is small.

In the bottom panel of Figure 11 we exhibit the parameters  $a_{0,2,3}$ , which inform us how the genus curve shape changes as we pass from real to redshift space. For the thick shell  $\Delta = 0.02$ , the parameters  $a_{0,3}$  are largely consistent with zero; however  $a_2$  marginally departs from zero for the thick shell  $\theta_G = 27.5$  arcmin case. In the thin shell case, we observe a more severe shape change in all three  $a_{0,2,3}$  parameters.

#### 4.4. Shot Noise and Galaxy Bias

There are two other systematic effects that must be considered when studying the genus: shot noise and galaxy bias. The former is an issue with sampling - as the number density of particles decreases, the point distribution becomes a successively poorer tracer of the underlying density field. The second effect will occur due to the variable relationship between samples of galaxies with different masses and the underlying matter power spectrum. It is not quite trivial to separate these two effects, as the application of mass cuts to the data will also decrease the number density of the catalog.

To study shot noise, we randomly cut dark matter particles from the HR4 sample. After constructing the genus



**Figure 11.** [Top panel] The amplitude coefficient  $a_1$  for  $\Delta g_{2D,RSD}$  for four sets of smoothing scales  $(\theta_G, \Delta)$ . The error bars are the error on the population mean of  $g_{2D,RSD}$ , estimated using our  $n = 8$  shells within the  $z = 0$  Horizon Run snapshot box. The black vertical line represents the linear approximation (36). The linear approximation correctly approximates the amplitude drop to within  $\sim 1.5\%$  for the smoothing scales considered in this work. [Bottom panel] Shape parameters  $a_{0,2,3}$  (blue, red, yellow) for the same  $(\theta_G, \Delta)$  smoothing scales. The shape parameters are approximately consistent with zero when using the thick shell  $\Delta = 0.02$ , although the  $a_2$  parameter exhibits  $\sim 1\%$  deviations from  $a_2 = 0$ .

of the cut sample, we compare it to the genus constructed using the full dark matter distribution. To analyze the effect of galaxy bias, we apply mass cuts to the galaxy catalog and compare the resulting genus to the full dark matter equivalent. We take mass cuts such that the random and mass cut catalogs contain the same number of points. The dark matter particle data that we use is an order of magnitude more dense than the galaxies, and so we expect it to accurately represent the underlying dark matter density field.

Removing galaxies from the sample will generically lead to an artificial increase in the number of structures, which become fragmented by the absence of data. The result will be an increase in the genus amplitude. Furthermore, randomly sampling the galaxies will produce a larger effect than a mass cut, as fragmentation of large overdensities is more likely to occur if we remove high-mass galaxies.

We define two residual genus functions  $\Delta g_{2D,sn}(\nu_A, \bar{d})$  and  $\Delta g_{2D,gb}(\nu_A, \bar{d})$

$$\Delta g_{2D,sn}(\nu_A, \bar{d}) = g_{2D,mat}(\nu_A, \bar{d}) - g_{2D,mat}(\nu_A, \bar{d}_0) \quad (39)$$

$$\Delta g_{2D,gb}(\nu_A, \bar{d}) = g_{2D,gal}(\nu_A, \bar{d}) - g_{2D,mat}(\nu_A, \bar{d}_0) \quad (40)$$

where  $g_{2D,mat}(\nu_A)$ ,  $g_{2D,gal}(\nu_A)$  are the genus curves extracted from the dark matter and mock galaxy  $z = 0$  snapshot data respectively, and  $\bar{d}$  is a measure of separation of galaxies on the sphere in arcminutes, defined as  $\bar{d} = (10800/\pi)\sqrt{4\pi/N_p}$  where  $N_p$  is the number of dark matter particles or galaxies in the cut catalog. Here,  $\bar{d}_0$  is the particle separation in the full dark matter sample. When we fit the Hermite expansion (16) to these quantities, the resulting Hermite coefficients  $a_{0-4}$  will provide a measure of how the genus is affected by sampling effects.

In Figure 12 we exhibit the amplitude coefficient  $a_1$  (top panel) and shape parameters  $a_{0,2,3}$  (bottom panel) as a function of  $\bar{d}/\theta_G$ . As throughout this work,  $a_{0,2,3}$  are represented by blue triangles, red stars, and yellow squares. The dashed/solid lines exhibit the effect of making random/mass cuts to the data. We have shifted the mass cut points linearly along the x-axis by a small value for clarity.

Clearly, the effect of sampling is large - the amplitude  $a_1$  increases monotonically with  $p/\theta_G$  when we take a mass cut, and at a greater rate when applying random cuts. This effect is potentially larger than any other considered in this work. As expected, randomly sampling the data produces a more significant effect on the amplitude than applying a mass cut. For a dense sample  $\bar{d}/\theta_G < 0.2$ , the  $a_1$  amplitude parameters for the random and mass cut catalogs are consistent with one another, within statistical fluctuations. Our choice of sampling is irrelevant, provided our catalog is sufficiently dense.

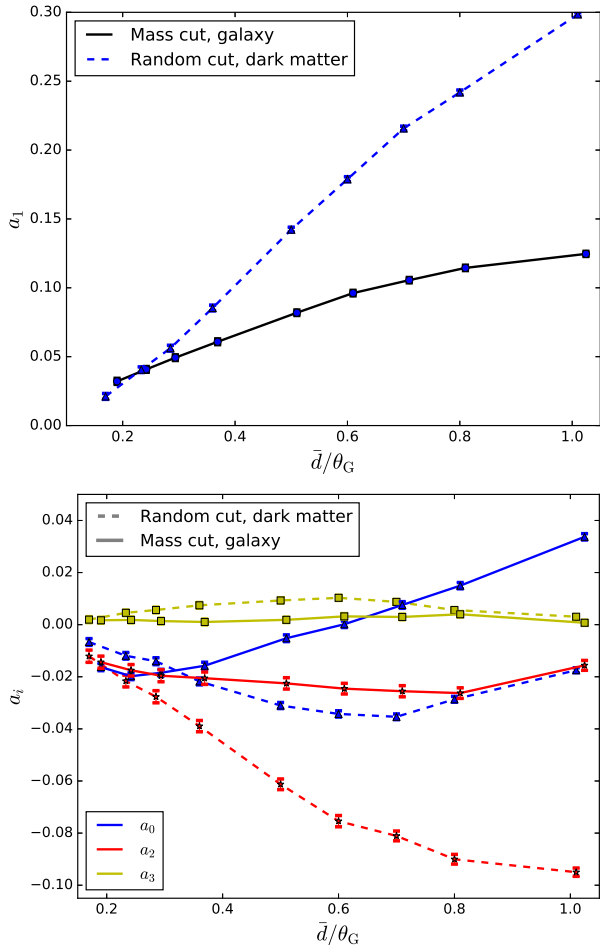
The shape of the genus curve is also modified as we apply increasingly large cuts to the data, most significantly  $a_2$ . Random sampling again has a more significant effect than applying a mass cut; however, this might be expected given that  $a_{0,1,2}$  are correlated. The increasingly negative  $a_2$  values observed in the bottom panel as we increase  $\bar{d}/\theta_G$  (red lines) suggest that the field becomes increasingly dominated by ‘meatball’ topology, i.e. a preponderance of isolated clusters. This agrees with our results in section 3.2 - as the correlation length of the sample increases, the genus has a tendency toward increasingly negative  $a_2$  values. We observe agreement in the shape parameters between the mass and randomly cut galaxy catalogs for  $\bar{d}/\theta_G < 0.2$ .

## 5. DISCUSSION

In this work we have performed an analysis of the properties of the two-dimensional genus of dark matter and mock galaxy catalogs projected onto tomographic redshift shells. We used the latest Horizon Run 4 dark matter simulation and considered effects of gravitational evolution, RSD, shot noise, galaxy bias, and finite size pixels. We draw the following conclusions

1. The finite size of pixels modifies the amplitude of the genus curve. However the effect is negligible if we ensure that the pixel size  $p$  satisfies  $p < \theta_G/3$ .
2. RSD will predominantly modify the genus amplitude. This effect is broadly consistent with linear





**Figure 12.** (Top panel) The amplitude coefficient  $a_1$  as a function of the ratio of the mean separation  $\bar{d}$  and smoothing scale  $\theta_G$ . The solid curve represents the galaxy data after successive mass cuts, and the dashed line is a randomly sampled catalog. The randomly sampled catalog exhibits a more extreme departure from the dark matter genus amplitude. (Bottom panel) The shape parameters  $a_{0,2,3}$ , correspond to blue triangles, red stars and yellow squares. The random and mass cut samples produce statistically equivalent genus curves for  $\bar{d}/\theta_G < 0.3$ .

predictions for smoothing scales  $R_G \simeq 10\text{Mpc}/h$  and thick redshift slices  $\Delta \simeq 0.02$ , although  $\sim 1\%$  discrepancies remain.

3. On small scales, the genus amplitude exhibits some evolution with redshift. We find that smoothing on scales  $R_G \simeq 10\text{Mpc}/h$  is sufficient to ensure that the genus amplitude evolves by  $\sim 1\%$  over the range  $0 < z < 1$
4. The effect of galaxy bias is sub-dominant provided the mean separation of galaxies on the sphere satisfies  $\bar{d}/\theta_G < 0.2$
5. The dominant departures in the genus shape from its Gaussian form can be described by the  $H_{0,2}$  Hermite polynomials. This agrees with analytic results in the mildly nonlinear regime (Matsubara 2000).
6. The genus amplitude and shape exhibit strong dependence on the number density of the galaxy sample. When comparing the genus at different red-

shifts, one should make mass cuts to ensure constant number density in each shell.

Cosmological information is encoded in the amplitude of the genus, so one must be particularly careful to avoid systematic effects when we compare its value in different redshift shells. The most important modifiers of this statistic are RSD and gravitational evolution, as both will explicitly depend upon the cosmological model adopted. If we wish to use the amplitude of the genus for parameter estimation, then these effects are effectively nuisance contributions. Nonlinear finger-of-God effects are mitigated by the stacking procedure that we are forced to adopt when we take  $\Delta > 0.02$ . We stress that, when dealing with galaxy data, we will not be free to specify the shell thickness - rather it must be larger than the photo-z redshift uncertainty.

We close with a brief discussion on the information content of the genus curve, and the different ways in which one can extract it. Both the shape and amplitude of the genus are related to ratios of cumulants of the density field. For example, the amplitude is a measure of  $\sigma_1^2/\sigma_0^2$  where  $\sigma_{0,1}$  are defined in equations (14) and (15). When the density field is smoothed on physical scales  $R_G \sim 10\text{Mpc}/h$  and the correct cosmology is adopted, the genus amplitude should be conserved as it is measured in successive redshift shells. Both the growth rate and linear bias are canceled when taking the ratio  $\sigma_1^2/\sigma_0^2$ , indicating that the amplitude is insensitive to these quantities. The shape of the angular power spectrum in each redshift shell (when smoothed on large scales) is also relatively insensitive to cosmology. Cosmological parameter dependence enters when we choose an incorrect cosmology, as this will introduce a systematic evolution in our choice of physical smoothing scale  $R_G$  and also in the surface area of the redshift shell. By choosing an incorrect  $R_G$  at different redshifts, one is effectively measuring the slope of the power spectrum at different points leading to a systematic increase or decrease of the genus amplitude. Hence, by comparing the genus curve at different redshifts, one can obtain cosmological parameter constraints by minimizing the evolution of the amplitude.

In a series of recent works Pogosyan et al. (2009); Gay et al. (2012); Codis et al. (2013), analytic modeling of the shape of the genus curve in the non-Gaussian low redshift universe has been undertaken via an expansion in the variance of the density field. Non-Gaussian shape modifications measure higher-order cumulants of the field. At first order, these are related to the three point functions cf. equations (29 – 31). There are two methods by which one can extract information from the non-Gaussian contributions. One is by measuring the higher-order cumulants and comparing them directly with theoretical predictions involving integrals of the Bispectrum, which can be calculated for both  $\Lambda\text{CDM}$  and modified gravity models. The second is by noticing that the non-Gaussian corrections are proportional to powers of  $\sigma_0$ , which will evolve with redshift via the growth factor  $D(z)$ . Hence by measuring the shape parameters at different redshifts, we can directly reconstruct  $D(z)$ .

Clearly, additional information is contained in the higher-order cumulants of the field, which can be extracted using the results of Codis et al. (2013). One advantage of using the non-Gaussian shape evolution of

the genus is that it would allow smoothing on smaller scales, increasing the statistical power of the method. The advantage of using the amplitude is that it provides a purely geometric test of the expansion rate. Because the amplitude and shape information is complementary, a combination of both will optimize constraints on cosmological parameters.

By calculating properties of the two-dimensional density field, we are losing information relative to its full three-dimensional counterpart. However, cosmological information is still contained within  $g_{2D}$  and in a forthcoming publication we will study the constraining power of this statistic when applied to galaxy data.

The authors thank the referee for his helpful comments and also the Korea Institute for Advanced Study for providing computing resources (KIAS Center for Advanced Computation Linux Cluster System) for this work.

This work was supported by the Supercomputing Center/Korea Institute of Science and Technology Information, with supercomputing resources including technical support (KSC-2013-G2-003) and the simulation data were transferred through a high-speed network provided by KREONET/GLORIAD.

Some of the results in this paper have been derived using the HEALPix (Gorski et al. 2005) package.

## REFERENCES

- Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., & Steward, L. 2013, arXiv:1311.5246, [ApJS.210.9(2014)]
- Blake, C., James, J. B., & Poole, G. B. 2014, MNRAS, 437, 2488
- Codis, S., Pichon, C., Pogosyan, D., Bernardeau, F., & Matsubara, T. 2013, Mon. Not. Roy. Astron. Soc., 435, 531
- Coles, P., Moscardini, L., Plionis, M., et al. 1993, MNRAS, 260, 572
- Coles, P., & Plionis, M. 1991, MNRAS, 250, 75
- Colley, W. N. 1997, ApJ, 489, 471
- Colley, W. N., Gott, J. R., Weinberg, D. H., Park, C., & Berlind, A. A. 2000, ApJ, 529, 795
- Gay, C., Pichon, C., & Pogosyan, D. 2012, Phys. Rev., D85, 023011
- Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ., 622, 759
- Gott, J. R., Colley, W. N., Park, C.-G., Park, C., & Mugnolo, C. 2007, MNRAS, 377, 1668
- Gott, J. R., Dickinson, M., & Melott, A. L. 1986, ApJ., 306, 341
- Gott, J. R., Park, C., Juskiewicz, R., et al. 1989
- Gott, J. R., Weinberg, D. H., & Melott, A. L. 1987, ApJ, 319, 1
- Gott, J. R., et al. 1988
- Hadwiger, H. 1957, Vorlesungen ber Inhalt, Oberfläche und Isoperimetrie (Grundlehren der mathematischen Wissenschaften: Springer)
- Hikage, C., Komatsu, E., & Matsubara, T. 2006, ApJ., 653, 11
- Hong, S. E., Park, C., & Kim, J. 2016, Astrophys. J., 823, 103
- Hoyle, F., Vogeley, M. S., & Gott, J. R. 2002, ApJ, 570, 44
- James, J. B. 2012, ApJ., 751, 40
- Jiang, C. Y., Jing, Y. P., Faltenbacher, A., Lin, W. P., & Li, C. 2008, Astrophys. J., 675, 1095
- Kim, J., Park, C., L’Huillier, B., & Hong, S. E. 2015, JKAS, 48, 213
- Kim, Y.-R., Choi, Y.-Y., Kim, S. S., et al. 2014, ApJS., 212, 22
- Laigle, C., et al. 2016, arXiv:1604.02350
- Lewis, A. 2013, Phys. Rev., D87, 103529
- Lewis, A., & Bridle, S. 2002, Phys. Rev., D66, 103511
- L’Huillier, B., Park, C., & Kim, J. 2014, New Astron., 30, 79
- Matsubara, T. 1994, arXiv:astro-ph/9501076
- Matsubara, T. 1996, ApJ, 457, 13
- . 2000, ArXiv Astrophysics e-prints, astro-ph/0006269
- . 2003, ApJ, 584, 1
- Matsubara, T., & Jain, B. 2001, ApJ., 552, L89
- Matsubara, T., & Suto, Y. 1996, ApJ., 460, 51
- Melott, A. L., Cohen, A. P., Hamilton, A. J. S., Gott, J. R., & Weinberg, D. H. 1989, ApJ, 345, 618
- Melott, A. L., Weinberg, D. H., & Gott, J. R. 1988, ApJ, 328, 50
- Park, C., & Gott, J. R. 1991, ApJ, 378, 457
- Park, C., Gott, J. R., & Choi, Y. J. 2001, ApJ, 553, 33
- Park, C., Gott, J. R., Melott, A. L., & Karachentsev, I. D. 1992, ApJ, 387, 1
- Park, C., Kim, J., & Gott, J. R. 2005, ApJ, 633, 1
- Park, C., & Kim, Y.-R. 2010, ApJ., 715, L185
- Pogosyan, D., Gay, C., & Pichon, C. 2009, Phys. Rev., D80, 081301, [Erratum: Phys. Rev.D81,129901(2010)]
- Ryden, B. S., Melott, A. L., Craig, D. A., et al. 1989, ApJ., 340, 647
- Schmalzing, J., & Gorski, K. M. 1998, MNRAS, 297, 355
- Schmalzing, J., Kerscher, M., & Buchert, T. 1996, Proc. Int. Sch. Phys. Fermi, 132, 281
- Speare, R., Gott, J. R., Kim, J., & Park, C. 2015, ApJ., 799, 176
- Wang, X., Chen, X., & Park, C. 2012, ApJ., 747, 48
- Wang, Y., Park, C., Xu, Y., Chen, X., & Kim, J. 2015, ApJ, 814, 6
- Weinberg, D. H., Gott, J. R., & Melott, A. L. 1987, ApJ, 321, 2
- Zunckel, C., Gott, J. R., & Lunnan, R. 2011, MNRAS, 412, 1402